

Prinule: komb. selekcia

$$EPE = E \left[ (y_{\text{new}} - x_{\text{new}}^T \hat{\beta})^2 \right] = \underbrace{\text{var}(y_{\text{new}} - x_{\text{new}}^T \hat{\beta})}_{\text{Variance}} + \underbrace{\left( E[y_{\text{new}} - x_{\text{new}}^T \hat{\beta}] \right)^2}_{\text{Bias}}$$

= (\*)

realita:  $y = f(x) + \varepsilon$

model:  $\hat{y} = x^T \beta + \varepsilon$

→ odhad:  $\hat{y} = x^T \hat{\beta}$  na zbl.  $y_1, \dots, y_n$   
 $x_1, \dots, x_n$

$$(*) = \text{Variance} + \text{Bias}^2 = \underbrace{\text{var}(y_{\text{new}})}_{\text{firna' variancia merani'}} + \underbrace{\text{var}(x_{\text{new}}^T \hat{\beta})}_{\text{variancia odhadov}} - \underbrace{2 \text{cov}(y_{\text{new}}, x_{\text{new}}^T \hat{\beta})}_0 + \text{Bias}^2$$

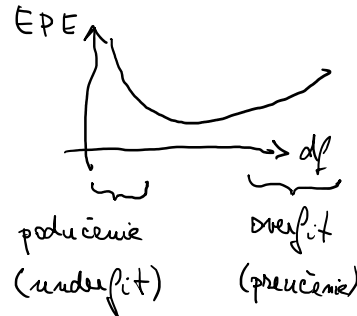
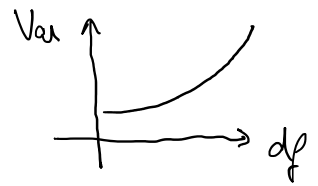
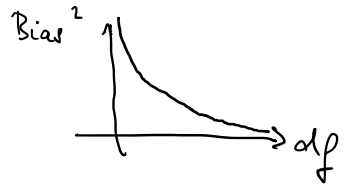
- bias-variance tradeoff

$$EPE = \text{var} + \text{bias}^2 \quad \square$$

- bias: vychylenie: model nie je schopny zachytit realitu ( $x^T \hat{\beta}$  je daleko od  $f(x)$ )
- variance: rozptyl: model sa muze na zblade dat rozhodnut o odhadoch parametrov (ma sada merani  $\Rightarrow$  vyrazne ine' predikcie)

$\square$  nalezdnost vbladom na  $y_1, \dots, y_n, y_{\text{new}}$

typicky: komplexnost modelu  $\uparrow$   $\Rightarrow$  variance  $\uparrow$   
 (# parametrov, degree of freedom: df)  $\Rightarrow$  bias  $\downarrow$



Kvalita modelu:

- EPE odhadneme pomocou testovacej vzorky

→ data rozdeline:   
 - trénovacia vzorka: Train   
 - testovacia vzorka: Test

→ odhadujeme parametre, vyberáme model   
 → kvalita finalneho modelu ( $MSE_{\text{test}}$ )